DOCUMENT RESUME

ED 441 017 TM 030 815

AUTHOR Schaffner, Monika; Burry-Stock, Judith A.; Cho, Gyu-Pan; Boney, Tracy; Hamilton, Gwen

TITLE What Do Kids Think When Their Teachers Grade?

PUB DATE 2000-04-00

NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).

PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150) -- Tests/Questionnaires (160)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS Elementary Secondary Education; Evaluation Methods; *Grades (Scholastic); Grading; *Student Attitudes; Student Evaluation; *Validity

ABSTRACT

Students' perceptions of assessment activities are the focus of this research. An instrument, Perceptions of Assessment of Teachers by Students, was developed in a primary version (kindergarten through grade 3) with a pictorial scale and a senior version (grades 4 through 12) with a 5-point Likert scale. Fifteen teachers returned the student inventories for 115 children in the primary group and 174 in the senior group (a 43% response rate). In this pilot study, the psychometric qualities of both versions appeared sound. There were discernible differences in grade levels in how students perceive their grades. Overall, students' sense of fairness and their perceived control over their own grades correlated significantly with teachers' self-reported perceptions of competence in assessment. However, analyses suggest that teachers are not asking students about what should be included in the grading process, and they indicate the importance of student perceptions of fairness. To make grading accurate, meaningful, and fair, putting the child into the teaching-testing-grading cycle is critical to understanding the validity of the assessment process. (Contains 1 figure, 4 tables, and 22 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 441 017

What Do Kids Think When Their Teachers Grade?

Monika Schäffner, Bowling Green State University, Judith A. Burry-Stock, Gyu-Pan Cho,

Tracy Boney, Gwen Hamilton, The University of Alabama

For more information please contact:

Monika Schäffner, Bowling Green State University, mschaff@bgnet.bgsu.edu

Judith A. Burry-Stock, The University of Alabama, jburry@bamaed.ua.edu

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

J. Burry-Stock

BEST COPY AVAILABLE

TM030815

Monika Schäffner, Bowling Green State University, Judith A. Burry-Stock, Gyu-Pan Cho, Tracy Boney, Gwen Hamilton, The University of Alabama

Assessment methods and theories have been the focus of much attention for over the past century. However, significant interest in assessment of classroom learning is a recent phenomenon. Educators have realized that real changes cannot take place in our schools without altering the way we assess student learning in the classroom (Torrance, 1995). The impact of assessment practices on learning is critical to student motivation. How and what a teacher grades affects how a student perceives him/herself on a day-to-day basis. If a student is doing well, they usually become confident learners and if not, the converse is often true. An issue that has escaped researchers and the like is what do the kids think about the assessment practices that are used to evaluate them. Their beliefs and understanding are important to their level of interest (Mitchell, 1992) and both intrinsic and extrinsic motivation (Husman & Lens, 1999) regarding school. Assessment activities, which ultimately impact grades, affect how students feel about specific content areas, courses in school, and may affect how they feel about school. Consequential validity of sorts (Messick; 1989; Popham,1997; Whittington, 1999) has not been used much in the context of classroom assessment, but it definitely applies. Students' perspectives on assessment activities are the focus of this research.

It is obvious, but not often acknowledged, that the outcome of teachers' assessment practices is student grades. The irony is that until recently, little research has been done on the topic of assessment practices, and even less has been done on students' perceptions of these practices (Plake; 1993, Popham;1997, Zhang & Bury-Stock,submitted). Who cares what kids think when their teacher grades? Anyone associated with education cares in some way what

happens when a teacher assigns a grade. These people cover the gamut from kids, to parents, teachers, administrators, politicians, and researchers. The purpose of this research was to investigate three aspects of the grading process: 1) how do students perceive the manner in which their teacher grade; 2) what is the relationship between the way teachers perceive their own grading practices and how students perceive them; and 3) what student perceived factors are related to their teachers' self-reported perceptions of competence in assessment?

It is imperative that we listen to children on how assessments influence their learning, their attitude toward education, and motivation. In order to work towards common goals, teachers and students need information on how well they are doing, and what they should change or improve to reach desired goals. Assessment directly influences student-teacher relationships, and how students are assessed inevitably affects how they are being taught. The teaching-testing-grading cycle is familiar to all students past 3rd or 4th grade, yet it is rarely fully comprehended and remains a mystery to many students, parents, and educators. However, students understand fair grading practices, and "fair assessment is one in which it is clear what will and will not be tested" (McMillan, 1997, p. 63).

If assessment is to be used as an instructional tool for achieving educational goals, students should be involved in this process. One could start by asking students what they think or feel about assessment, because one of the key purposes of assessment is to ask important questions about student learning. Listening to students' perceptions of their teachers' assessment practices is a rare phenomenon. Yet, we know that when students are involved in assessment through student-led conferences in portfolio assessment, they are more likely to assume responsibility for their own goals and learning (Conderman, Hatcher & Ikan, 1998).

When Spage (1996) asked her students what they perceived as the most effective tool for measuring their learning, she learned that her students felt a gap between what they had learned and what the typical "test" indicated they had learned. Based on their responses she tailored her assessment tools around their needs, which resulted in a cooperative atmosphere in which the children felt more empowered and more motivated to learn. Different teachers have different assessment practices, and children have to determine how each teacher is going to assess them, because their grades depend on the assessment process. To date, literature and formal research on students' perceptions of their teachers' assessment practices is scarce.

Instrument Development

The instrument used in this study is named "Perceptions of Assessment of Teachers by Students" (PATS) and was developed as a Primary Version (K – 3rd grade), and a Senior Version (4th – 12th grade). Asking students two open-ended questions generated items on the instruments: 1) "What do you think about tests and why?" and 2) "Why do you think that your teacher gives you grades?" Answers to these open-ended questions were obtained from students at the intermediate and high school grade levels from a nearby school system, and were used to generate scale items. A group of professionals, who were trained in assessment of classroom learning, comprised the team that constructed and revised both PATS. The instrument 4 - 12 was first pilot-tested on 80 high school students in a large city in a southeastern state. Item data were collected and the instruments were revised.

The PATS (K – 3) Primary Version has 15 items with a three-point pictorial scale. Items were selected from the original instrument and rewritten to adjust for the reading and cognitive levels. The Primary PATS is to be read to the students, following the reading of the items on

their copy. Students circle the picture that best represents their feeling about the statement on the following scale: ☹ for disagreement, 😐 for neutral, and ☺ for agreement.

The PATS 4 - 12 (Senior) version has 55 items with a five-point Likert-type scale, ranging on a continuum from "Never" to "Always," e.g., "Never 1 / 2 / 3 / 4 / 5 Always." These instruments measure students' self-reported experiences regarding classroom grades and assessment practices such as: fairness issues, curricular issues, and relevance issues. Figure 1 shows selected items for PATS (primary) and for PATS (senior).

Insert Figure 1 here

Methodology

Teachers who participated in a 1998 statewide study of perceived competence in classroom assessment, as measured by the Assessment Practices Inventory (API) (Zhang & Burry-Stock, 1994), were asked to participate again in this study. These teachers were contacted by telephone to obtain permission for measuring their students' perceptions of their grading practices. Thirty-five teachers agreed to participate, and were mailed the Perception of Assessment Practices of Teachers by Students (PATS) inventory (Burry-Stock, et al., 1999; Schäffner, et al., 1999). The instruments were mailed to teachers, who administered them to their students in the spring of 1999. Fifteen teachers returned the student inventories for 115 children in the K – 3rd grade group and 174 students in the 4th – 12th grade group, giving us a 43% response rate. Table 1 depicts the various sampling categories.

Insert Table 1 here

In the K – 3rd grade group 53% were boys and 47% were girls. The 4th – 12th grade sample consisted of 45% boys and 55% girls. African American students were 43% of the total group, which reflects the demographics in the public schools in this state. 24% identified themselves as Caucasians, and 30% of the students preferred to leave this part blank. The majority of students were in the 5th, 6th, 9th, and 11th grades with smaller numbers in 7th, 8th, 10th and 12th.

Data Analysis

The psychometric qualities were tested for both versions of PATS. An internal consistency reliability coefficient (Cronbach's alpha) for the total was equal to .93 for the Senior PATS. A Cronbach's alpha equal to .73 was obtained for the K – 3rd instrument. An exploratory factor analysis was done on the Senior version using a principal components method with a varimax rotation that revealed 45.83 % of the variability. A confirmatory factor analysis was done on the Senior version, using the LISREL 8.3 program (Jöreskog & Sörbom, 1993). This was done to justify the formation of subscales to be used in the regression analysis to answer research question 3. These factors otherwise known as subscales are shown in Table 2, with the number of items that make up each subscale, the Cronbach alpha reliability coefficient for each subscale, and the Goodness-of-Fit index for the measurement model. The subscale titles with the reliability coefficients are as follows: 1) fairness issues, 87; 2) teacher's job, .85; 3) internal locus of control, .83; 4) attitudes towards grades and grading, .61; 5) student input in grading, .48; and 6) essay, there is no reliability coefficient, because it is not appropriate for a one item subscale. Fairness has always been an issue as has the teachers' job or duty to evaluate students (Scriven, 1991). Internal locus of control may be linked to internal motivation (Husman & Lens, 1999),

and attitude may be a combination of both internal and external motivation. Student input has recently been important to some teachers especially those involved in a more active/constructivist type of teaching (Burry-Stock & Yager, 1999). For whatever reason the essay test remained a separate factor with an item number of one, which assisted in establishing the confirmatory factor analysis. Essay items are the link between paper-pencil items and performance items; perhaps students do know this and they may not know how to view them. Statistics for the confirmatory factor analysis show a Chi-square with 55 degrees of freedom equal to 89.332 where alpha equals .01 and the root mean square (RMS) of .059 is close to the required .05, which indicates that the measurement model fit the data for a relatively small sample of 174. A larger sample would have shown more stability with the statistics and possibly a better fit. However, the data fit the model well enough to allow further investigations.

Insert Table 2 here

A five-factor solution was obtained for the junior version by using an exploratory factor analysis principal component solution and a varimax rotation, which explained 58% of the total variance. Factors, subscales, in order of their contribution to the factor analysis are: 1) Internal locus of Control, 2) Fairness Issues, 3) Positive Grade Impact, 4) Teacher's Job, and 5) Negative Grade Impact. Interpretation for the subscale titles is the same as they are above; however, the factor structures were not stable and did not provide a good simple structure. Hence no further analysis was done for the Primary PATS.

Reliability and validity evidence was sufficient for further analysis of the Secondary PATS so as to address the first research question. 1) Are there grade-level differences in

students' perceptions of their teachers' grading practices? A One-way analysis of variance (ANOVA) was statistically significant at alpha equals .001, ($F_{9,162} = 9.104$, $p < .001$). Therefore, post-hoc paired comparisons were run to investigate the differences between the grade levels. Table 3 details the results of these analyses.

Insert Table 3 here

Significant differences were found among grade levels using Tukey's honestly significant difference test (HSD). Differences occurred between the: 5th and 11th, and 8th and 9th grades ($p < .05$); 4th and 6th, and 5th and 9th grades ($p < .01$); 6th and 9th, and 6th and 11th grades ($p < .001$). Grade levels where there were no differences had small samples, so it is possible that there are differences between all grade levels. Clearly, students at different grade levels perceive the assessment process differently. The greatest differences occur between the intermediate grades, 4 – 6, and the high school levels, 9 – 11, where grades actually have an impact on the child's future career decisions.

The second research question is, 2) Is there a relationship between teachers' self-reported perceptions of competence in assessment and their students' perceptions of their grading practices? Since the PATS was a follow-up study, data were available from the prior year on how teachers perceived their skill in using various assessment practices. The instrument used is called the Assessment Practice Inventory (API) (Zhang & Burry-Stock, 1994, 2000). The API was written to reflect the Classroom Assessment Standards (American Federation of Teachers, National Council on Measurement in Education, & National Education Association 1990). Items on a Likert-type scale from one to five were used to measure teachers' perceived assessment skills

regarding: the development of instruments, the use of various types of instruments, the analysis of items, and ethics. The simple correlation for the Senior PATS and the API is .36, $p < .001$ which is moderate, and indicates that as the students felt more positive about the way their teachers assessed them, and the teachers felt more positive about their skill in using assessment practices. No significant relationship was found in the primary version.

The third research question was: 3) Which student perceived factors are related to their teachers' self-reported perceptions of competence in assessment? In order to answer this question a backward multiple regression analysis was run using the total teacher scores on the API as the dependent variable and the subscales of the PATS senior version as the independent variables. The results are shown in Table 4.

Insert Table 4 here

The full model $R^2 = .318$ and the R-value is .564. A .05 level of significance was chosen for the variables to remain in the model. Variables in order of deletion are: the Teacher's Job subscale and the Attitude scale. Remaining variables in the model were Fairness Issues, Internal Locus of Control, Student Input, and Essay as statistically significant variables. The R^2 change was .002 with a final R^2 of .316. These results indicate that students' sense of fairness and their perceived control over their own grades correlates significantly with their teachers' self-reported perceptions of competence in assessment, which is why it is important to listen to children about grading and grades. It is also interesting that the correlation between student input and their teachers' competence in assessment is negatively related, which shows that teachers are not yet amenable to allowing students to give input on what they will be assessed. The item on the essay test remained in the model, because most students responded that their teachers do not give any essay tests.

Conclusion

The psychometric qualities of both versions of both instruments appear sound in this pilot study, but further research should be done using larger sample sizes. There are discernable differences between grade levels in how students perceive their grades. The factor structures, subscales, although tenuous, do provide insight into subdimesions of how students view assessment practices. Subscales for the Senior PATS are: Fairness Issues, Teacher's Job, Internal Locus of Control, Attitude, Student Input, and Essay. In a prior study involving the same teachers and the Assessment Practice Inventory (API) (Zhang & Burry-Stock, 1994), which is a measure of a teacher's perception of how they view their assessment skills, a final model R^2 of .3186 was obtained in a multiple regression analysis. Four of the subscales remained in the model and were statistically significant. In order of their contribution to the model they are: Student Input (an inverse relation); Internal Locus of Control; Fairness; and Essay (an inverse relationship). From these analyses it appears that we are not asking students about what should be included in the assessment process; that internal locus of control can be interpreted as internal motivation is important (Husman & Lens, 1999); fairness in the grading process is important ; and somehow essay tests are viewed differently from other assessments.

Examining classroom assessment procedures and outcomes brings a new meaning to the word "validity." Since grades are an accumulation of assessments, it seems to us that validity of the various assessments, and the grading process are important. Originally validity was thought to be a mathematical process established by psychometricians. According to Whittington, (1999, p. 15), "For teachers, this mathematical way of thinking feels removed from the day-to-day realities of instruction." Since ultimately all aspects of validity are related to construct validity, which lead to "value implications" and "social consequences" (Messick, 1989), we need to

examine the use of the term validity for classroom assessment. As was so aptly stated by Messick, (1989, p. 59) in defining validity "Furthermore, derivation for the term 'value' from the old French *valior* meaning 'to be worth,' applies as well to modern uses of 'valid,' as references to the functional worth of the testing."

Since we hold students accountable for learning and teachers for teaching, we are looking at an underlying assumption that students will learn if properly instructed, yet we are not ready to allow students to have a say on what they will be graded (Newman, King, & Rigdon, 1997). Assessment practices are extremely important, because they ultimately lead to evaluation of students' achievement and are reported to many audiences. Thus, many values and social consequences are made based upon these assessments and grades. Often students' academic self esteem, parents' perception of their children, and other educators' judgment is measured by how well students do in school. Incorporating their perceptions of the grading process should be helpful in providing teachers and other educators with information to strengthen the assessment process. Understanding how this process can be accurate, meaningful, and fair is extremely important to teaching and learning. Putting the child into the teaching-testing-grading cycle is critical to understanding the validity of the assessment process.

References

American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *Standards for teacher competence in educational assessment of students.* Washington, DC: National Council on Measurement in Education.

Bergin, D.A. (1999). Influences on classroom interest. *Educational Psychologist,34(2),* 86-98.

Burry-Stock, J.A., Schäffner, M., Cho, G. P.Boney, T., Hamilton, G.E. McNeese, M. (1999). Perception of teachers grading practices by students (PATS) 4 -12. Tuscaloosa, AL: The University of Alabama.

Burry-Stock, J.A., Schäffner, M., Cho, G. P.(submitted). Making a difference in teachers perceived skill competence: the use of a standards driven curriculum in teaching classroom assessment. *Applied Measurement In Education.*

Burry-Stock, J.A., Schäffner, M., Cho, G. P., & McNeese, M., Zhang, Z. (1999). Using a standards-driven curriculum in teaching classroom assessment: making a difference in teachers' perceived skill competency. A paper presented at the Annual Meeting of the American Educational Research Association (AERA) in Montreal, Canada, April 1999.

Burry-Stock, J.A. & Yager, R. E. (1999). How long does it take to become a constructivist expert science teacher. Paper presented at the national conference for the American Educational Research Association (AERA), Montreal, CA.

Conderman, G.,Hatcher R.E., & Ikan, P.A. (1998). Why student-led conferences work. *Kappa Delta Pi Record, 34 (4)*, 132-134.

Husman, J. & Lens, W. (1999). Influences on classroom interest. *Educational Psychologist,34* (2), 113-125.

Jöreskog, K. & Sörbom, D. (1993). *Liserl 8: Structural Equation Modeling with the SIMPLIS Command Language*. Chicago, IL: Scientific Software International.

McMillan, J.H. (1997). *Classroom Assessment: Principles and Practice for Effective Instruction*. Needham Heights, MA: Allyn & Bacon.

Messick, S. (1989). Validity in *Educational Measurement Third Edition*, Samuel Messick (Ed.). New York, New York: American Council on Education and Macmillan Publishing Company a Division of Macmillan, Inc.,13-104.

Mitchell, R. (1992). *Testing for Learning: How New Approaches to Evaluation Can Improve American Schools*. New York, NY: The Free Press.

Newman, F.M., King, M.B., & Rigdon, M. (1997). Accountability and school performance: implications from restructuring schools. *Harvard Educational Review, 67 (1)*, 47-74.

Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher, 6(1)*, 21-27.

Popham, W. J. (1997). Consequential validity: right concern-wrong concept. *Educational Measurement: Issues and Practice, 16*(2), 9-13.

Schäffner, M., Burry-Stock, J.A.,Cho, G. P.Boney, T., Hamilton, G.E. McNeese, M. (1999). Perception of teachers grading practices by students (PATS) - 3. Tuscaloosa, AL: The University of Alabama.

Scriven, M (1991). Teacher evaluation models project memorandum. Kalamazoo: The Center for Research on Educational Accountability and Teacher Evaluation. Western Michigan University

Spage, L. (1996). Listening to children on issues of assessment. *Teaching and Change, 3* (4), 359-410.

Torrance, H. (Ed.). ((1995). *Evaluating Authentic Assessment*. Buckingham: England: Open University Press.

Whittington, D. (1999) Making room for values and fairness: teaching reliability and validity in the classroom context. *Educational Measurement*, 18, (1), 14 – 22.

Zhang, Z. & Burry-Stock, J.A. (1994). Assessment practice inventory (API). Tuscaloosa, Al: The University of Alabama

Zhang, Z., & Burry-Stock, J. A. (submitted, 2000). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*.

Perception of Assessment of Teachers by Students (PATS), Primary

1. I know where I stand in class. ☹ 😐 ☺
2. My teacher knows when I do not understand. ☹ 😐 ☺
3. My teacher lets me know when I have done well. ☹ 😐 ☺
4. My teacher answers my questions. ☹ 😐 ☺
5. I get bad grades because I misbehave in class. ☹ 😐 ☺

Perception of Assessment of Teachers by Students (PATS), Senior

1. My teacher grades me fairly. Never 1 / 2 / 3 / 4 / 5 Always
2. My teacher grades me on many different things.
3. My grades show what I have learned.
4. I get bad grades because I misbehave in class.
5. I have a say in deciding on what I will be tested.
6. My teacher likes to give essay tests.

Figure 1: Selected items representing each factor, first for PATS (Primary) then for PATS (Senior).

Table 1: Sample Characteristics of Primary and Senior PATS

Primary (K – 3)		Senior (4 – 12)					
Gender		Gender		Race/Ethnicity		Grade Level in School	
Boys	61	Boys	79	African American	75	Grade 4	14
Girls	54	Girls	95	Caucasian	41	5	30
Total	115	Total	174	Hispanic	2	6	43
				Asian	2	7	6
				Other	2	8	7
				Missing	52	9	30
						10	5
						11	36
						12	3

Table 2: Confirmatory factor analysis with subscales of the Senior PATS

Subscale	#of Items	Cronbach's Alpha	Goodness-of-fit Statistics
Fairness Issues	14	.87	$X^2_{55} = 89.332$ (P > .01)
Teachers' Job	13	.85	RMS = .059
Int. Loc. of Control	15	.83	GFI = .927
Attitude	7	.61	N = 174
Student Input	5	.48	
Essay	1	NA	

Table 3: ANOVA and Tukey's HSD for the differences among means for the different grade levels on the Senior PATS

ANOVA	Sum of Squares	df	Mean Square	F	Significance
Between Groups	50058.945	9	5562.105	9.104	.000
Within Groups	100200.69	164	610.980		
Total	150259.63	173			

Comparisons among Means with Tukey's HSD

Grade levels in school	Mean Difference	Std. Error	Significance	95% Confidence Interval	
				Lower Bound	Upper Bound
4th and 6th grade	-32.3744	8.6779	.007	-59.8287	-4.9201
5th and 9th grade	25.6667	6.3822	.002	5.4755	45.8578
5th and 11th grade	21.5833	6.1105	.015	2.2518	40.9149
6th and 9th grade	43.6744	5.8800	.000	25.0718	62.2770
6th and 11th grade	39.5911	5.5839	.000	21.9252	57.2569
8th and 9th grade	33.2857	10.3754	.044	0.4612	66.1102

Table 4: Backward Linear Regression with the Assessment Practice Inventory (API) Total as the Dependent Variable and the Senior PATS Subscales as Predictor Variables

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	R	R^2
	B	Std.Error	Beta				
1 (Constant)	122.733	31.263		3.926	.000	.564	.31
FAIRNESS	1.109	0.683	0.169	1.624	.106		
TEACHJOB	0.123	0.718	0.017	0.171	.864		
INTLOCNT	1.774	0.597	0.274	2.971	.003		
STDINPUT	-4.810	1.170	-0.290	-4.112	.000		
ATTITUDE	0.741	1.087	0.052	0.681	.497		
ESSAY	-8.771	3.236	-0.185	-2.711	.007		
2 (Constant)	123.265	31.018		3.974	.000	.564	.31
FAIRNESS	1.169	0.585	0.178	1.999	.047		
INTLOCNT	1.803	0.572	0.278	3.155	.002		
STDINPUT	-4.795	1.163	-0.289	-4.123	.000		
ATTITUDE	0.745	1.083	0.052	0.688	.492		
ESSAY	-8.768	3.226	-0.185	-2.718	.007		
3 (Constant)	131.788	28.392		4.642	.000	.562	.31
FAIRNESS	1.279	0.652	0.195	2.277	.024		
INTLOCNT	1.894	0.555	0.293	3.413	.001		
STDINPUT	-4.921	1.147	-0.297	-4.292	.000		
ESSAY	-8.842	3.219	-0.187	-2.747	.007		